Appendix C

Performance Characteristic Evaluation

Appendix C PERFORMANCE CHARACTERISTIC EVALUATION

To determine the method precision (i.e., measurement error within a site), we evaluated two data sets from SNARL, one from the Leviathan Mine study and another from the Upper Truckee River, and a large database (CalEDAS) from the California Department of Fish & Game containing CSBP data. The Leviathan Mine data set included a total of seven metrics, which were calculated from 54 sites (Table 1). On the other hand, the Upper Truckee River included a total of 15 metrics, which were calculated from 18 sites (Table 2). Where there were common metrics, the data was combined, and several metrics were calculated from a total of 72 sites (Table 3). The data set using the CSBP method was significantly larger (approximately 360 sites) and was much more widely distributed than the SNARL data; however, details on the exact site distribution across the state were not provided.

An analysis of variance (ANOVA) was conducted to compare the variability among replicates at each site. From the mean squared error (MSE), we calculated the root mean square error (RMSE), which can be used to compare precision between metrics, and the coefficient of variability (CV), which can be used to compare precision among metrics. The RMSE provides an estimate of the standard deviation of a population of observations; however, it is scale dependent, and therefore metrics that are on different scales cannot be directly compared. CV, on the other hand, is a unit-less measure calculated by dividing the RMSE by the mean of the dependent variable, which allows for direct comparison among means and indices. Because the CV takes into account the within site variability relative to the sample mean, it was chosen to be the better indicator of precision when comparing the two methods.

Tables 1 and 2 list ANOVA results of SNARL data from the Leviathan Mine dataset and the Upper Truckee River dataset, respectively. Unfortunately, the same metrics were not calculated for both studies; therefore, in our attempt to combine the datasets, the number of observations is not consistent among the different metrics (i.e., N = 18, N = 72)(Table 3). Table 4 lists the among season variability for data collected in the Upper Truckee River study using the SNARL method.

Table 1. ANOVA results of SNARL Leviathan Mine data (N = 54)

Metric	*MS Error*	*RMSE*	*Mean*	*CV*
Species Richness	12.05	3.47	23.52	14.76
EPT Taxa	2.98	1.73	9.00	19.18
Density (ind./m^2)	212000000.00	14551.03	15653.30	92.96
%Chironomidae	0.01	0.07	0.35	20.88
Ratio EPT/Chironomidae	1.38	1.17	1.34	87.24
Hilsenhoff Biotic Index	0.11	0.33	4.52	7.20
Dominance	0.01	0.08	0.40	20.50

Table 2. ANOVA results of SNARL Upper Truckee River data (N =18)

Metric	*MS Error*	*RMSE*	*Mean*	*CV*
Species Richness*	14.17	3.76	27.09	13.90
EPT TAXA*	3.42	1.85	11.10	16.67
No. EPHEMEROPTERA TAXA	1.07	1.03	6.77	15.26
No. PLECOPTERA TAXA	1.58	1.26	4.33	28.99
No. TRICHOPTERA TAXA	1.34	1.16	5.73	20.22
No. of CHIRONOMIDAE	5.98	2.45	11.22	21.80
No. of Individuals	49468.56	222.42	593.24	37.49
DENSITY* (#/m^2) @30x30 cm area	170000000.	13054.01	17948.72	72.73
%EPT	0.01	0.09	0.63	15.00
%CHIRONOMIDAE*	0.00	0.07	0.31	22.44
Chiro Richness / Chiro Density	0.00	0.00	0.01	62.39
Hilsenhoff Biotic Index*	0.13	0.36	4.26	8.46
%TOLERANT TAXA (7-8-9-10)	0.00	0.05	0.11	47.70
INTOLERANT TAXA (0-1-2)	4.66	2.16	14.41	14.98
DOMINANCE*	0.01	0.08	0.36	21.52
%FILTER-FEEDERS	0.00	0.05	0.08	60.62

Table 3. ANOVA results of combined SNARL data (N =18)

Metric	*MS Error*	*RMSE*	*Mean*	*CV*
Species Richness	19.76	4.44	36.80	12.08
EPT TAXA	4.61	2.15	16.83	12.75
No. EPHEMEROPTERA TAXA	1.07	1.03	6.77	15.26
No. PLECOPTERA TAXA	1.58	1.26	4.33	28.99
No. TRICHOPTERA TAXA	1.34	1.16	5.73	20.22
No. of CHIRONOMIDAE	5.98	2.45	11.22	21.80
No. of Individuals	49468.56	222.42	593.24	37.49
DENSITY (no./m^2) @30x30 cm area	61354347.0	7832.90	24197.34	32.37
%EPT	0.01	0.09	0.63	15.00
%CHIRONOMIDAE	0.01	0.08	0.20	38.18
Chiro Richness / Chiro Density	0.00	0.00	0.01	62.39
Hilsenhoff Biotic Index	0.19	0.44	3.58	12.31
%TOLERANT TAXA (7-8-9-10)	0.00	0.05	0.11	47.70
INTOLERANT TAXA (0-1-2)	4.66	2.16	14.41	14.98
DOMINANCE	0.00	0.06	0.25	25.46
%FILTER-FEEDERS	0.00	0.05	0.08	60.62

Table 5 lists the metrics used to describe the characteristics of the benthic macroinvertebrate communities sampled according to each method. It should be noted that the metrics listed in the table are not part of a biological index for either method, and the metrics calculated for each study does not necessarily remain consistent. Therefore, the suite of metrics listed in this table is not intended to be indicative of the analyses performed for each study.

Table 4. Among season CVs for SNARL Upper Truckee River data.

Metric	*Spring 95*	*Spring 97*	*Fall 98*	*Spring 99*	*Fall 99*
Species Richness	13.79	22.72	13.12	15.02	14.05
EPT TAXA	28.21	20.71	19.02	21.24	14.48
No. of Individuals	50.92	27.83	29.49	36.79	23.59
Density (ind / m^2)	50.92	27.83	35.61	47.55	72.93
%Chironomidae	19.58	24.74	21.13	13.80	21.56
Ratio EPT/Chironomidae	60.05	60.37	84.01	119.34	59.55
Hilsenhoff Biotic Index	4.86	11.50	8.66	6.63	4.90
Dominance	35.64	22.46	19.97	19.02	19.53

* N = 72 for these metrics.

Table 5. Metrics used to by each method to describe characteristics of Benthic Macroinvertebrate communities.

Metric	**CSBP**	**SNARL**	**Metric**	**CSBP**	**SNARL**
Taxa Richness	X	[1]	% Hydropsychidae	X	
EPT Taxa	X	X	% Baetidae	X	
Ephemeroptera Taxa	X	X	% Dominant Taxa	X	X
Plecoptera Taxa	X	X	% Collectors	X	
Trichoptera Taxa	X	X	% Filterers	X	X
Chironomidae Taxa		X	% Scrapers	X	
EPT Index (%)	X	X	% Predators	X	
Sensitive EPT Index	X	[2]	% Shredders	X	
Shannon Diversity Index	X		Density		X
Hilsenhoff Biotic Index		X	Estimated Abundance	X	X
Tolerance Value	X		Ratio EPT/ Chironimdae		X
% Intolerant Organisms	X		Chironomidae Richness/ Chironomidae Density		X
% Tolerant Organisms	X	X			

Footnotes:

1 Species Richness

2 Number of Intolerant Taxa

Table 6 lists the ANOVA results of the CSBP dataset. Table 7 shows the ANOVA results of both datasets and can be used to compare precision estimates between methods. Because the CSBP data set contained a much larger number of observations (N =300), we decided to standardize the number of observations and compare the results to see if observation size had any significant effect on differences in precision (Table 8). Furthermore, we standardized the number of replicates between the two datasets to see if replicate size had any effect on differences in precision (Table 9).

Table 6. ANOVA results of CSBP data (N = 300)*

Metric	MS Error	RMSE	MEAN	CV
TAXA RICHNESS	10.33	3.21	16.72	19.23
EPT Taxa	2.54	1.59	6.45	24.71
Ephemeroptera Taxa	0.53	0.72	2.97	24.44
Plecoptera Taxa	1.19	1.09	2.83	38.54
Trichoptera Taxa	1.01	1.00	2.82	35.65
Chironomid Taxa	0.46	0.68	2.46	27.47
%EPT	127.56	11.29	42.21	26.76
%Chironomidae	71.46	8.45	19.99	42.29
Hilsenhoff Biotic Index	0.63	0.79	5.77	13.70
% Tolerant Taxa	126.30	11.24	22.37	50.23
Intolerant Taxa	0.89	0.94	2.99	31.49
Dominance	138.83	11.78	43.45	27.12

* For plecoptera taxa metric, N = 168

Table 7. Comparison of ANOVA results between CSBP and SNARL methods.

Metric	CSBP			SNARL			% Difference
	RMSE	MEAN	CV	RMSE	MEAN	CV	
Richness	3.21	16.72	19.23	3.76	27.09	13.9	5.4
EPT Taxa	1.59	6.45	24.71	1.85	11.1	16.67	8.04
Ephemeroptera Taxa	0.72	2.97	24.44	1.03	6.77	15.26	9.18
Plecoptera Taxa	1.09	2.83	38.54	1.26	4.33	28.99	9.55
Trichoptera Taxa	1	2.82	35.65	1.16	5.73	20.22	15.43
Chironomid Taxa	0.68	2.46	27.47	2.45	11.22	21.8	5.67
%EPT	11.29	42.21	26.76	9.5	63.32	15	11.76
%Chironomidae	8.45	19.99	42.29	6.99	31.15	22.44	19.85
Hilsenhoff Biotic Index	0.79	5.77	13.7	0.36	4.26	8.46	5.24
% Tolerant Taxa	11.24	22.37	50.23	5.4	11.32	47.7	2.53
Intolerant Taxa	0.94	2.99	31.49	2.16	14.41	14.98	16.51
Dominance	11.78	43.45	27.12	7.78	36.16	21.52	5.6

Table 8. Comparison of precision estimates between CSBP and SNARL methods where population size is consistent (N = 18)

	CSBP				SNARL			
Metric	MS Error	RMSE	Mean	CV	MS Error	RMSE	Mean	CV
EPT Taxa	1.63	1.28	5.87	21.75	4.61	2.15	16.83	12.75
Ephemeroptera Taxa	0.31	0.56	2.65	21.19	1.07	1.03	6.77	15.26
Plecoptera Taxa	0.63	0.79	0.91	87.45	1.58	1.26	4.33	28.99
Trichoptera Taxa	0.67	0.82	2.04	40.08	1.34	1.16	5.73	20.22
Chironomidae Taxa	0.43	0.65	2.20	29.62	5.98	2.45	11.22	21.80
%EPT	102.42	10.12	47.88	21.14	0.90	9.50	63.32	15.00
%Chironomidae	76.66	8.76	20.10	43.55	0.60	7.74	20.27	22.44
Hilsenhoff Biotic Index	0.43	0.65	4.74	13.74	0.19	0.44	3.58	12.31

Table 9. Comparison of precision estimates between CSBP and SNARL methods where replicate size is consistent (replicates = 3).

	CSBP				SNARL			
Metric	MS Error	RMSE	Mean	CV	MS Error	RMSE	Mean	CV
EPT Taxa	2.54	1.59	6.45	24.71	3.14	1.77	10.66	16.60
Ephemeroptera Taxa	0.53	0.72	2.97	24.44	0.81	0.90	6.70	13.47
Plecoptera Taxa	1.19	1.09	2.83	38.54	1.11	1.05	4.28	24.64
Trichoptera Taxa	1.01	1.00	2.82	35.65	1.24	1.11	5.69	19.59
Chironomidae Taxa	0.46	0.68	2.46	27.47	4.67	2.16	11.00	19.64
%EPT	127.56	11.29	42.21	26.76	0.49	7.03	64.10	10.97
%Chironomidae	71.46	8.45	19.99	42.29	0.53	7.31	32.36	22.58
Hilsenhoff Biotic Index	0.63	0.79	5.77	13.70	0.17	0.41	4.31	9.45

<u>Case Example Defining Method Performance Characteristics</u>

While developing a statewide network for biomonitoring and bioassessment using macroinvertebrate data, Florida Department of Environmental Protection (DEP) rigorously examined performance characteristics of their collection and assessment methods in order to provide better overall quality assurance of their biomonitoring program and to provide defensible and appropriate assessments of the state's surface waters (Barbour et al. 1996b, c). This case example was summarized from Chapter 4 - Performance-Based Methods System in *Rapid Bioassessment Protocols for Use in Wadeable Streams and Rivers* (Barbour et al. 1999).

Characterizing Sampling Error (Method Precision on a Population of Reference Sites): A total of 56 reference sites were sampled in the Peninsula bioregion. The Florida Stream Condition Index (SCI) score could range from a minimum of 7 to a theoretical maximum of 31 based on the component metric scores. However, in the Peninsula, reference site SCI scores generally ranged between 21 and 31. A mean SCI score of 27.6 was observed with a CV of 12.0%.

Determining Method and Index Sensitivity: Distribution of SCI scores of the 56 reference sites showed that the 5th percentile was a score of 20. Thus, 95% of Peninsula reference sites had a score >20. Accuracy of the method, using known stressed sites, indicated that approximately 80% of the test sites had SCI scores ≤ 20. In other words, a stressed site would be assessed as impaired 80% of the time using the collection method in the Peninsula bioregion in the summer, and an impairment criterion of the 5th percentile of reference sites.

Determination of Method Bias and Relative Sensitivity in Different Site Classes: A comparative analysis of precision, sensitivity, and ultimately bias, was performed for the Florida DEP method and the SCI index. The mean SCI score in the Panhandle bioregion, during the same summer index period, was 26.3 with a CV = 12.8% based on 16 reference sites. Comparing this CV to the one reported for the Peninsula above, it is apparent that the precision of this method in the Panhandle was similar to that observed in the Peninsula bioregion. On the other hand, the 5th percentile of the Panhandle reference sites was an SCI score of 17, such that actual sensitivity of the method in the Panhandle was slightly lower than in the Peninsula bioregion. An impaired site would be assessed as such only 50% of the time in the Panhandle bioregion during the summer as opposed to 80% of the time in the Peninsula bioregion during the same index period.